Largest-fit selection of random sizes under a sum constraint: comparisons by weak convergence

Frans A. Boshuizen
Econometric Institute, Erasmus Universiteit Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Robert P. Kertz
School of Mathematics,
Georgia Institute of Technology,
Atlanta GA 30332, USA

March 20, 1995

Short title: Selection of random sizes.

Proofs should be sent to:

Professor R.P. Kertz
School of Mathematics
Georgia Institute of Technology
Atlanta GA 30332, USA

Abstract. Let $Y_1, Y_2, \cdots$ be a sequence of positive, independent, identically distributed random variables with distribution function $G(x)$ and denote the order statistics of $Y_1, \ldots, Y_n$ by $Y_{1,n} \leq \ldots \leq Y_{n,n}$. For positive numbers $(d_n)_{n\geq 1}$, the largest-fit off-line counting random variable $\hat{N}^e(d_n)$ is defined by $\hat{N}^e(d_n) = \max\{j : 1 \leq j \leq n \text{ and } Y_{n,n} + \cdots + Y_{n-j+1,n} \leq d_n\}$ if this set is nonempty and $= 0$ otherwise. In a bin-packing context, the r.v. $\hat{N}^e(d_n)$ is the maximal number of objects with largest possible sizes (chosen from n objects sampled from distribution G) that can be packed into a bin of capacity d_n.

In the on-line situation the n objects with sizes $Y_1, \ldots, Y_n$ arrive one-by-one, and a decision maker must either accept or reject each object immediately, without recall. A 'good' on-line selection policy selects as few objects as possible with sizes as large as possible.

In this paper, quantitative comparisons are given between the off-line count $\hat{N}^e(d_n)$ and on-line counts $N_n^\tau(d_n)$ associated with 'good' policies τ. Specifically, for such policies τ, under appropriate conditions on the distribution function $G(x)$ and the constants $(d_n)_{n\geq 1}$, we find sequences of positive constants $(\beta_n)_{n\geq 1}$, $(\Delta_n)_{n\geq 1}$ and $(\Delta'_n)_{n\geq 1}$ such that

$$\left(\Delta_n \left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1\right), \Delta'_n \left(\frac{N_n^\tau(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (W, W') \text{ as } n \to \infty,$$

for some nondegenerate r.v.'s W and W'.

1 Introduction

Let $Y_1, Y_2, \cdots$ be i.i.d. positive r.v.'s with distribution function $G(x)$. For positive numbers $(d_n)_{n\geq 1}$, the *largest fit counting r.v.* $\hat{N}^e(d_n) = \hat{N}^e_n(d_n)$ is defined by

$$\hat{N}^e(d_n) := \max\{j : 1 \leq j \leq n \text{ and } Y_{n,n} + \cdots + Y_{n-j+1,n} \leq d_n\}$$

if this set is nonempty and $= 0$ otherwise. In the bin-packing context, the r.v. $\hat{N}^e(d_n)$ is the maximal number of objects with largest possible sizes that can be packed into a bin of capacity d_n, chosen from n objects with random sizes sampled from distribution G. This is the largest count obtainable by a 'prophet' or individual using an off-line strategy, that is, a strategy which uses knowledge of all sizes in the sample, without any order restrictions on the sample values. Let τ^e denote this largest-fit off-line strategy. Note that, under this strategy, if the j-th largest item has been packed, and the $(j+1)$-st largest item does not fit in with the j largest one's, then the strategy stops with j items and does not proceed to pack items with smaller sizes. One also has the representation

$$\hat{N}^e(d_n) = \min\{j : 1 \leq j \leq n \text{ and } Y_{n,n} + \cdots + Y_{n-j+1,n} > d_n\} - 1$$

if this set is nonempty and $= n$ otherwise.

Define a *policy* τ to be a sequence of stopping times $(\tau_j)_{j\geq 1}$ with respect to $Y_1, Y_2, \cdots$ with $1 \leq \tau_1 < \tau_2 < \cdots$. For n objects, the counting r.v. associated with policy τ is defined by $N^\tau_n := \sum_{j\geq 1} I(\tau_j \leq n)$. For any policy $\tau = (\tau_j)_{j\geq 1}$ satisfying the sum constraint $\sum_{j\geq 1} Y_{\tau_j} I(\tau_j \leq n) \leq d_n$, the r.v. N^τ_n has interpretation in a bin-packing context as the number of objects that are packed into a bin of size d_n, chosen sequentially under policy τ, without recall, as n objects with sizes sampled from distribution G appear in a given order. This N^τ_n is an 'on-line' count, a count under an on-line policy.

In this paper, quantitative comparisons are given between the off-line count $\hat{N}^e(d_n)$ and the on-line counts $N^\tau_n(d_n)$ associated with 'good' strategies τ for $(Y_j)_{j\geq 1}$. Intuitively, a 'good' on-line policy for this setting is one which selects only items with large sizes and stops when the total-size sum reaches d_n. Difficulties arise if one seeks to translate these features into an on-line, expectation-based optimization problem. For example, the choice of optimization

$$\inf\left\{ EN^\tau_n : \tau \text{ is a policy for } (Y_j)_{j\geq 1} \text{ such that } \sum_{j\geq 1} Y_{\tau_j} I(\tau_j \leq n) > d_n \right\}$$

rewards choice of large values of the Y_j's, but rules out policies $\tau = (\tau_j)_{j\geq 1}$ with $\sum_{j\geq 1} Y_{\tau_j} I(\tau_j \leq n) \leq d_n$ on any part of the underlying probability space; and the choice of an optimization with a penalty feature in the optimization

$$\inf\{E\hat{N}_n^\tau(d_n) : \tau \text{ is a policy for } (Y_j)_{j\geq 1}\}$$

where $\hat{N}_n^\tau(d_n) = \sum_{j=1}^n I(\tau_j \leq n)$ if $\sum_{j\geq 1} Y_{\tau_j} I(\tau_j \leq n) \geq d_n$ and $= n$ if $\sum_{j\geq 1} Y_{\tau_j} I(\tau_j \leq n) < d_n$, again rewards choice of large values of the Y_j's, but imposes too great a penalty for the situation $\sum_{j\geq 1} Y_{\tau_j} I(\tau_j \leq n) < d_n$. Explicit solution of such on-line expectation-based optimization problems also appears to be more difficult than for the comparable smallest-fit analogues in Coffman, Flatto and Weber [5].

To place this problem, consider the following comparison. A manufacturer of a liquid chemical product anticipates production of an amount d_n of the product during the next business cycle. At the beginning of the period, the manufacturer knows there will be n customers (one order per customer) during the period. The manufacturer can fill orders partially if limited by her production level. The company requires that the manufacturer service as few orders as possible, with these being of largest possible order size, and should have total sum of order sizes attaining the anticipated production level d_n. If the manufacturer has knowledge of the sizes of all customer orders at the beginning of the business period, then she will service largest orders first and thereby fill $\hat{N}^e(d_n)$ full orders. If the manufacturer does not know the sizes of all customer orders at the beginning of the business period, but sees customer orders one-by-one as they arrive, and must either accept or reject each order immediately, without recall, then she must establish some criteria and choose a 'good' policy for handling orders under this criteria.

In this paper, the objective is to obtain asymptotic joint distributional comparisons of the off-line count $\hat{N}^e(d_n)$ and on-line counts $N_n^{\hat{\tau}}$ for 'good' policies $\hat{\tau}$. In analogy with the policies described in the smallest-fit problem discussed in Boshuizen and Kertz [2], the specific policy in the analysis and the comparison criteria are defined as follows.

For positive constants $(\rho_n)_{n\geq 1}$ and $(d_n)_{n\geq 1}$ and n objects with sizes $Y_1, \ldots, Y_n$ the *stopped threshold policy* $\hat{s}_n$ *with horizon* n accepts objects of size $\geq \rho_n$ and selects these only when the sum of the sizes of the selected objects is $\leq d_n$; if the sum of the sizes $\geq \rho_n$ up to the present is $> d_n$, then the present object and all future objects are rejected. Thus $\hat{s}_n$ is the policy of stopping times $(\tau_j)_{j\geq 1}$ for $(Y_j)_{j\geq 1}$ defined by $\tau_1 = \min\{1 \leq j \leq n : Y_j \leq d_n \text{ and } Y_j \geq \rho_n\}$ if this set is nonempty and $= \infty$ otherwise, and for $k = 2, 3, \cdots$, $\tau_k = \min\{\tau_{k-1} < j \leq n : Y_j \geq \rho_n \text{ and } \sum_{i=1}^j Y_i I(Y_i \geq \rho_n)) \leq d_n\}$ if $\tau_{k-1} < \infty$ and this set is nonempty, and $= \infty$

otherwise. The counting r.v. associated with this policy at horizon n, denoted $N^{\hat{s}_n}(d_n) = N_n^{\hat{s}_n}(\rho_n, d_n) := N_n^{\hat{s}_n}$ has representation $N^{\hat{s}_n}(d_n) = \sum_{i=1}^{\hat{\nu}_n(d_n)} I(Y_i \geq \rho_n)$ where $\hat{\nu}_n(d_n) = \max\{j : 1 \leq j \leq n$ and $\sum_{i=1}^{j} Y_i I(Y_i \geq \rho_n) \leq d_n\}$ if this set is nonempty and $= 0$ otherwise. A sequence of policies $(\hat{\tau}_n)_{n\geq 1}$ with $\hat{\tau}_n = (\hat{\tau}_{n,j})_{j\geq 1}$, satisfying the sum constraints $\sum_{j\geq 1} Y_{\hat{\tau}_{n,j}} I(\hat{\tau}_{n,j} \leq n) \leq d_n$ is said to be a **consistent approximator** of the largest-fit off-line strategy τ^e if there exists positive constants $(\beta_n)_{n\geq 1}$, $(\Delta_n)_{n\geq 1}$ and $(\Delta'_n)_{n\geq 1}$ for which

$$\left(\Delta_n\left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1\right), \Delta'_n\left(\frac{N^{\hat{\tau}_n}(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (W, W') \tag{1}$$

as $n \to \infty$, for some nondegenerate r.v.'s W and W'. In Theorems 2.1 and 3.1, it is shown that the sequence of stopped threshold policies $(\hat{s}_n)_{n\geq 1}$ is a consistent approximator of τ^e; and consequences of these results are given in Corollaries 2.6 and 3.4.

Observe that one could combine the theorems of this paper with the smallest-fit results of Boshuizen and Kertz [2], if one allows negative X_i's in [2]. However, the present arrangement is more natural for the given applications with positive r.v.'s. In fact, the theorems in this paper complement but do not duplicate the results in [2]. Note, in particular, the interesting variations for asymptotic behaviors that occur for $G(x)$ in Case II for maxima, as indentified in Theorems 2.1 and 3.1. The reader is also refered to the paper by Bruss and Robertson [3] where expressions for the asymptotic behavior of $E\hat{N}^e(d_n)$ are obtained (under weaker assumption on the distribution $G(x)$).

Throughout the paper the following notation is used. For a nondecreasing function h on a subset S of $\mathbb{R}$, the left-continuous inverse of h is defined by $h^{\leftarrow}(s) = \inf\{x \in S : h(x) \geq s\}$. For a distribution function G, $r_G = \sup\{y : G(y) < 1\}$ denotes the right end point of the support of G. The notations $o_P(1)$ and $O_P(1)$ are used to denote sequences of random variables which are respectively converging to zero in probability and bounded above and below by a finite constant uniformly for all n large. For two r.v.'s X and Y, $X \stackrel{\mathrm{d}}{=} Y$ if the distributions of X and Y are same. For a real number x, $\lceil x \rceil$ denotes the smallest integer greater than x, and for two sequences $(m_n)_{n\geq 1}$ and $(l_n)_{n\geq 1}$ we write $m_n \approx l_n$ if $\lim_{n\to\infty} m_n/l_n = 1$.

2 Normal convergence

In this section, settings are considered where appropriate normalizations of the counting r.v.'s $\hat{N}^e(d_n)$ and $N^{\hat{s}_n}(d_n)$ converge weakly (as $n \to \infty$) to a random pair of the form $(W_1, W_2 + (W_3 \wedge 0))$ where (W_1, W_2, W_3) has a (singular) multivariate normal distribution. The settings considered are $G(x)$ in Cases I, II with index $\alpha \geq 2$, and III for maxima (see the Appendix for background results on extreme value theory). Case II with index $0 < \alpha < 2$ will be treated in Section 3. Recall that for $G(x)$ in Case II for maxima, with index $\alpha > 0$, the inverse function $G^{\leftarrow}$ has representation $G^{\leftarrow}(1-s) = s^{-a}L(s)$ for $L(s)$ slowly varying at zero and $a = 1/\alpha$; and for $G(x)$ in Case III for maxima, with index $\alpha > 0$, the inverse function $G^{\leftarrow}$ has representation $G^{\leftarrow}(1-s) = r_G - s^{-a}L(s)$ for $L(s)$ slowly varying at zero and $a = -1/\alpha$. For G in Case I for maxima the function $c(s)$ is defined by $c(s) = s^{-1}\int_{1-s}^{1}(1-u)dG^{\leftarrow}(u)$.

For r.v.'s Y_i with $EY_i < \infty$, and for a given sequence of positive constants $(d_n)_{n\geq 1}$, define constants $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ by

$$d_n = n\int_0^{\beta_n/n} G^{\leftarrow}(1-s)ds \text{ and } \rho_n = G^{\leftarrow}(1-(\beta_n/n)). \tag{2}$$

In the following settings of $G(x)$ in the Cases I,II and III for maxima, it is assumed that the constants $(\beta_n)_{n\geq 1}$ satisfy $\beta_n/n \to 0$ as $n \to \infty$. This implies the following for the normalizing constants of Theorem 2.1: the location parameter $\beta_n << n$ and scaling parameters α_n, $\gamma_n << n$ as $n \to \infty$. This ensures that the total number of sizes selected goes to infinity, but stays small when compared to the horizon n, as $n \to \infty$, for the off-line largest fit strategy τ^e and for the on-line stopped threshold policy $\hat{s}_n$. Note, however, that in the Cases I, II with $0 < a \leq 1$ and III, $\hat{\nu}_n(d_n)/n \to 1$ in probability as $n \to \infty$; in this sense, it is taking closer and closer to n observations for the stopped threshold policy $\hat{s}_n$ to achieve total sum of sizes near the level d_n as $n \to \infty$. Also, under the assumptions in the theorems of this paper, it follows that $\rho_n << d_n$, so the policies $\hat{s}_n$ are feasible, for all n large.

It is also assumed that

$$(i) \qquad \beta_n^{1/2}\left(\frac{1-G(\rho_n-)}{\beta_n/n} - 1\right) \to 0 \text{ as } n \to \infty, \text{ and} \tag{3}$$

and

(ii) $\quad \alpha_n \to \infty \quad$ as $\quad n \to \infty$ where

$$\alpha_n \approx \begin{cases} \beta_n^{1/2}c(\beta_n/n)/G^{\leftarrow}(1-(\beta_n/n)) & \text{for } G(x) \text{ in Case I} \\ \beta_n^{1/2}(r_G - G^{\leftarrow}(1-(\beta_n/n)))/r_G & \text{for } G(x) \text{ in Case III with } r_G > 0 \end{cases}$$

Condition $(3i)$ is imposed to ensure that the location parameters in the convergence of the two counting r.v. sequences of Theorem 2.1 can be taken to be the same β_n. In the case that $G(x)$ is a continuous d.f., it is immediate that $1 - G(\rho_n-) = \beta_n/n$; so condition $(3i)$ holds in this case. For d.f. $G(x)$ in Case I,II or III for maxima, it follows that

$$1 \leq (1-G(\rho_n-))/(\beta_n/n) \leq (1-G(\rho_n-))/(1-G(\rho_n)) \to 1 \text{ as } n \to \infty;$$

so the limit in $(3i)$ can be thought of as a condition on the rate of convergence of $(1 - G(\rho_n-))/(\beta_n/n)$ to 1. In the Appendix we give an example of a d.f. $G(x)$ in Case II for maxima with $a = 1/3$ for which condition $(3i)$ fails. If d.f. $G(x)$ is in Case II with $a = 1/2$, one may replace $(3i)$ by the weaker assumption

$$\left(\frac{\beta_n}{\log \beta_n}\right)^{1/2}\left(\frac{1-G(\rho_n-)}{\beta_n/n} - 1\right) \to 0 \text{ as } n \to \infty \tag{3b}$$

whenever $(3i)$ is assumed for this case in this section.

Condition $(3ii)$ is imposed to ensure that the scaling parameters in the convergence of the off-line counting r.v. sequence in Theorem 2.1 are sufficiently large (see Example 2.5 (i) and (ii)). For $G(x)$ in Case II, no additional condition is needed in Theorem 2.1 to ensure that $\alpha_n \to \infty$ in this case.

Theorem 2.1 *Let $G(x)$ be in Case I, II with $0 < a \leq 1/2$, or III for maxima with $G(0-) = 0$ and $0 < r_G \leq \infty$, and let the positive constants $(d_n)_{n\geq 1}$, $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ satisfy (2). Assume that $d_n/n \to 0$ and $\beta_n \to \infty$ as $n \to \infty$ and (3) holds. Then there exist positive constants $(\alpha_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq 1}$ for which*

$$\left(\alpha_n^{-1}\left(\hat{N}^e(d_n) - \beta_n\right), \gamma_n^{-1}\left(N^{\hat{s}_n}(d_n) - \beta_n\right)\right) \Rightarrow \left(\hat{\mathcal{N}}^e, \hat{\mathcal{N}}^s\right)$$

where $\left(\hat{\mathcal{N}}^e, \hat{\mathcal{N}}^s\right) = (W_1, W_2 + (W_3 \wedge 0))$ and $\mathbf{W} = (W_1, W_2, W_3)$ is $N(\mathbf{0}, \Sigma_{\mathbf{W}})$-distributed.

For $G(x)$ in Case I, the constants $(\alpha_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq 1}$ satisfy

$$\alpha_n \approx \beta_n^{1/2} c(\beta_n/n)/G^{\leftarrow}(1 - (\beta_n/n)) \text{ and } \gamma_n \approx \beta_n^{1/2},$$

and $\Sigma_{\mathbf{W}} = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}.$

For $G(x)$ in Case II, the constants $(\alpha_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq 1}$ satisfy

$$\alpha_n = \gamma_n \approx \begin{cases} \beta_n^{1/2} & \text{for} \quad 0 < a < 1/2 \\ (\beta_n \log \beta_n)^{1/2} & \text{for} \quad a = 1/2 \end{cases}$$

and $\Sigma_{\mathbf{W}} = \begin{pmatrix} K_a^2 & \frac{-a}{1-a} & \frac{a}{1-2a} \\ \frac{-a}{1-a} & 1 & -1 \\ \frac{a}{1-2a} & -1 & \frac{(1-a)^2}{1-2a} \end{pmatrix}$ *for* $0 < a < 1/2$, *and* $= \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 1/4 \end{pmatrix}$ *for* $a = 1/2$

where $K_a^2 = \frac{2a^2}{(1-a)(1-2a)}$.

For $G(x)$ in Case III, the constants $(\alpha_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq 1}$ satisfy

$$\alpha_n \approx \beta_n^{1/2}\left(r_G - G^{\leftarrow}(1 - (\beta_n/n))\right)/r_G \text{ and } \gamma_n \approx \beta_n^{1/2}$$

and $\Sigma_{\mathbf{W}} = \begin{pmatrix} K_a^2 & \frac{a}{1-a} & \frac{-a}{1-a} \\ \frac{a}{1-a} & 1 & -1 \\ \frac{-a}{1-a} & -1 & 1 \end{pmatrix}$.

In fact, for Cases I and III, $W_3 = -W_2$*; and for Case II,* $W_3 = (1-a)(W_1 - W_2)$ *for* $0 < a < 1/2$ *and* $W_3 = (1/2)W_1$ *for* $a = 1/2$.

Note that it is not true in general that $N^{\hat{s}_n}(d_n) \le \hat{N}^e(d_n)$. However, Theorem 2.1 implies that $\lim_{n\to\infty} P(N^{\hat{s}_n}(d_n) \le \hat{N}^e(d_n)) = 1$ in Cases I and III, since

$$P(N^{\hat{s}_n}(d_n) \le \hat{N}^e(d_n)) = P(\beta_n^{-1/2}(N^{\hat{s}_n}(d_n) - \beta_n) \le \delta_n \alpha_n^{-1}(\hat{N}^e(d_n) - \beta_n)) \to P(\hat{\mathcal{N}}^s \le 0) = 1$$

(for specific $\delta_n \to 0$) in Cases I and III. In Case II with $0 < a \le 1/2$, Theorem 2.1 implies

$$P(N^{\hat{s}_n}(d_n) \le \hat{N}^e(d_n)) = P(\alpha_n^{-1}(N^{\hat{s}_n}(d_n) - \beta_n) \le \alpha_n^{-1}(\hat{N}^e(d_n) - \beta_n)) \to P(\hat{\mathcal{N}}^s \le \hat{\mathcal{N}}^e) = 1/2$$

The proof of Theorem 2.1 is based on a Brownian bridge approximation to the uniform empirical process. Before the theorem is proved, some lemmas concerning the Brownian bridge approximation are given. In the sequel we work on a probability space $(\Omega, \mathcal{A}, P)$ constructed by Csörgő et al. [6] carrying an infinite sequence $U_1, U_2, \cdots$ of i.i.d. r.v.'s uniformly distributed on $(0,1)$ and a sequence of Brownian bridges $U_n(s)$, $0 \le s \le 1$, $n = 1, 2, \cdots$ such that for the uniform emprical process $\alpha_n(s) = n^{1/2}(E_n(s) - s)$, $0 \le s \le 1$, where $E_n(s) = n^{-1}\sum_{i=1}^n I(U_i \le s)$

$$\sup_{1/n \le s \le 1-1/n} n^\nu \frac{|\alpha_n(s) - U_n(s)|}{(s(1-s))^{(1/2)-\nu}} = O_P(1) \text{ as } n \to \infty, \tag{4}$$

where ν is any fixed number such that $0 \le \nu < 1/4$. This can be assumed without loss of generality. Note that for an G-distributed r.v. Y_i we have $Y_i \stackrel{\mathrm{d}}{=} G^{\leftarrow}(U_i)$, so on this space we use $Y_i = G^{\leftarrow}(U_i)$ for $i = 1, 2, \cdots$. Recall that the constants $(\beta_n)_{n\ge1}$ and $(\rho_n)_{n\ge1}$ are defined in equation (2). The proofs of the following two lemmas use standard arguments, for example, found in the papers by Csörgő and Mason [9], Csörgő, Haeusler and Mason [11], and Lo [12].

Lemma 2.2 *Let* $G(x)$ *be in Case I, II with index* $\alpha \ge 2$*, or III for maxima with* $G(0-) = 0$*,* $0 < r_G \le \infty$*, let* $(\beta_n)_{n\ge1}$ *satisfy* $\beta_n/n \to 0$ *and* $\beta_n \to \infty$*, and let* $(k_n)_{n\ge1}$ *be any sequence of positive integers such that* $k_n \approx \beta_n$*. Then*

$$\frac{\sum_{i=1}^{k_n} Y_{n-i+1,n} - n\int_0^{k_n/n} G^{\leftarrow}(1-s)ds}{n^{1/2}A_n^e(k_n/n)} = -\frac{\int_{1-(\beta_n/n)}^{1-(1/n)} U_n(s)dG^{\leftarrow}(s)}{A_n^e(\beta_n/n)} + o_P(1)$$

$$\text{where for } 0 < t < 1,\ A_n^e(t) = \begin{cases} t^{1/2}c(t) & \text{in Case I} \\ t^{(1/2)-a}L(t) & \text{in Case II with } 0 < a < 1/2 \\ & \quad \text{and in Case III} \\ \left(\int_{1/n}^{t} u^{-1}L^2(u)du\right)^{1/2} & \text{in Case II with } a = 1/2 \end{cases}.$$

Lemma 2.3 *Let $G(x)$ be in Case I, II with index $\alpha \geq 2$, or III for maxima with $G(0-) = 0$ and $0 < r_G \leq \infty$, and $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ satisfy (2) with $\beta_n/n \to 0$ and $\beta_n \to \infty$ and let $\tau_n := 1 - G(\rho_n-)$ for $n \geq 1$. Then*

$$(n\tau_n)^{-1/2}\left(\sum_{i=1}^{n} I(Y_i \geq \rho_n) - n\tau_n\right) = -\tau_n^{-1/2}U_n(1-\tau_n) + o_P(1)$$

and

$$\frac{\sum_{i=1}^n Y_i I(Y_i \geq \rho_n) - n\int_0^{\tau_n} G^{\leftarrow}(1-s)ds}{n^{1/2}A_n^s(\tau_n)}$$

$$= \begin{cases} -\tau_n^{-1/2}U_n(1-\tau_n) + o_P(1) \text{ in Cases I and III} \\ -\tau_n^{-1/2}U_n(1-\tau_n) - \dfrac{\int_{1-\tau_n}^{1-(1/n)} U_n(s)dG^{\leftarrow}(s)}{A_n^s(\tau_n)} + o_P(1) \\ \qquad\qquad \text{in Case II with } 0 < a < 1/2 \\ -\dfrac{\int_{1-\tau_n}^{1-(1/n)} U_n(s)dG^{\leftarrow}(s)}{A_n^s(\tau_n)} + o_P(1) \text{ in Case II with } a = 1/2 \end{cases}$$

$$\text{where for } 0 < t < 1,\ A_n^s(t) = \begin{cases} t^{1/2}G^{\leftarrow}(1-t) & \text{in Cases I, II with } 0 < a < 1/2 \text{ and III} \\ \left(\int_{1/n}^{t} u^{-1}L^2(u)du\right)^{1/2} & \text{in Case II with } a = 1/2 \end{cases}.$$

If, in addition, condition $(3i)$ holds as a hypothesis in Lemma 2.3, then the conclusions hold with $\tau_n = 1 - G(\rho_n-)$ replaced by β_n/n. Note that the norming constants $A_n^e(t)$ and $A_n^s(t)$ coincide in Case II but differ in Cases I and III, just as α_n and γ_n in Theorem 2.1.

The next lemma is also needed in the proof of Theorem 2.1. The proof of the lemma uses the representations of $G^{\leftarrow}(1-s)$ in each of the Cases I, II and III, and results on slowly varying functions. (See the Appendix.)

Lemma 2.4 *Let $(d_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ be the sequences of positive constants defined in (2), let $k_n = \lceil \alpha_n\mu + \beta_n \rceil$, $-\infty < \mu < \infty$, and $j_n = \lceil n(1 + \gamma_n\beta_n^{-1}z) \rceil$, $-\infty < z < 0$, and let the constants $(\alpha_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq 1}$ be given as in Theorem 2.1. Assume condition $(3ii)$ holds. Then*

$$\frac{d_n - n\int_0^{k_n/n} G^{\leftarrow}(1-s)ds}{n^{1/2}A_n^e(k_n/n)} = -\mu + o(1) \text{ and}$$

$$\frac{d_n - (j_n d_n/n)}{n^{1/2}A_n^s(\beta_n/n)} = \begin{cases} -z + o(1) & \text{in Case I and III} \\ -z/(1-a) + o(1) & \text{in Case II with } 0 < a \leq 1/2 \end{cases}$$

where the functions $A_n^e(t)$ and $A_n^s(t)$ are given as in Lemmas 2.2 and 2.3, respectively.

Proof of Theorem 2.1. Consider first $G(x)$ in Case II with $0 < a < 1/2$. For the convergence analysis of $N^{\hat{s}_n}(d_n)$, observe that

$$\frac{N^{\hat{s}_n}(d_n) - \beta_n}{\beta_n^{1/2}} = \frac{\sum_{i=1}^{\hat{\nu}_n(d_n)} (I(Y_i \geq \rho_n) - (1 - G(\rho_n -)))}{\beta_n^{1/2}} + \beta_n^{1/2}\left(\frac{\hat{\nu}_n(d_n)}{n} - 1\right) + o_P(1)$$

where $\hat{\nu}_n(d_n) = \max\{j : 1 \leq j \leq n \text{ and } \sum_{i=1}^{j} Y_i I(Y_i \geq \rho_n) \leq d_n\}$ if this set is nonempty and $= 0$ otherwise; and obtain the following representations associated with this sum by using the argument in the proof of the Doeblin-Ascombe Central Limit Theorem as given in the book by Chow and Teicher [4, Theorem 1, page 317], the result that $\hat{\nu}_n(d_n)/n \to 1$ in probability, and Lemma 2.3:

$$\beta_n^{-1/2} \sum_{i=1}^{\hat{\nu}_n(d_n)} (I(Y_i \geq \rho_n) - (1 - G(\rho_n -))) = -(\beta_n/n)^{-1/2} U_n(1 - (\beta_n/n)) + o_P(1) \text{ and}$$

$$\frac{\sum_{i=1}^{j_n} Y_i I(Y_i \geq \rho_n) - (j_n d_n)/n}{n^{1/2}(\beta_n/n)^{1/2} G^{\leftarrow}(1 - (\beta_n/n))} = -(\beta_n/n)^{-1/2} U_n(1-(\beta_n/n)) - \frac{\int_{1-(\beta_n/n)}^{1-(1/n)} U_n(s) dG^{\leftarrow}(s)}{(\beta_n/n)^{1/2} G^{\leftarrow}(1 - (\beta_n/n))} + o_P(1)$$

where $j_n := \lceil n(1 + z_2/\beta_n^{1/2})\rceil$ with $-\infty < z_2 < 0$. Also define integers $k_n := \lceil \alpha_n \mu + \beta_n \rceil$ where the constants $(\alpha_n)_{n\geq 1}$ are defined in the statement of the theorem, and use the representations above and the convergence results from Lemma 2.4 to obtain the weak convergence

$$\begin{aligned}
&P\left(\frac{\hat{N}^e(d_n) - \beta_n}{\alpha_n} \leq \mu; \frac{\sum_{i=1}^{\hat{\nu}_n(d_n)} (I(Y_i \geq \rho_n) - (1 - G(\rho_n -)))}{\beta_n^{1/2}} < z_1; \beta_n^{1/2}\left(\frac{\hat{\nu}_n(d_n)}{n} - 1\right) \leq z_2\right) \\
&= P\left(\sum_{i=1}^{k_n} Y_{n-i+1,n} > d_n; \frac{\sum_{i=1}^{\hat{\nu}_n(d_n)} (I(Y_i \geq \rho_n) - (1 - G(\rho_n)))}{\beta_n^{1/2}} < z_1; \sum_{i=1}^{j_n} Y_i I(Y_i \geq \rho_n) > d_n\right) \\
&= P\left(\frac{\int_{1-(\beta_n/n)}^{1-(1/n)} U_n(s) dG^{\leftarrow}(s)}{A_n^e(\beta_n/n)} < \mu; -(\beta_n/n)^{-1/2} U_n(1 - (\beta_n/n)) < z_1;\right. \\
&\qquad \left.(1-a)\left((\beta_n/n)^{-1/2} U_n(1 - (\beta_n/n)) + \frac{\int_{1-(\beta_n/n)}^{1-(1/n)} U_n(s) dG^{\leftarrow}(s)}{A_n^s(\beta_n/n)}\right) < z_2\right) + o(1) \\
&= P(W_1 < \mu;\ W_2 < z_1;\ W_3 < z_2) + o(1)
\end{aligned}$$

where $\mathbf{W} = (W_1, W_2, W_3) \overset{\mathrm{d}}{=} N(0, \Sigma_{\mathbf{W}})$ and the covariance matrix $\Sigma_{\mathbf{W}}$ is defined in the assertion of the theorem.

Hence, for $-\infty < \mu < \infty$ and $-\infty < \nu < \infty$,

$$P\left(\frac{\hat{N}^e(d_n) - \beta_n}{\alpha_n} \leq \mu; \frac{N^{\hat{s}_n}(d_n) - \beta_n}{\beta_n^{1/2}} \leq \nu\right) = P(W_1 \leq \mu; W_2 + (W_3 \wedge 0) \leq \nu) + o(1)$$

and the theorem is proved for $G(x)$ in Case II with $0 < a < 1/2$.

For the proof of the theorem in all other cases, use straightforward modifications of this argument, together with the Lemmas 2.2, 2.3 and 2.4. $\square$

Example 2.5 (*i*) For Exponential(1)-distribution G (so $G(x)$ is in Case I for maxima), and positive constants $(\beta_n)_{n\geq 1}$ with $\beta_n \to \infty$, $\beta_n/n \to 0$ and $\beta_n^{1/2}/\log(n/\beta_n) \to \infty$, and for $d_n = \beta_n - \beta_n \log(\beta_n/n)$ and $\rho_n = -\log(\beta_n/n)$, Theorem 2.1 gives

$$\left(\beta_n^{1/2} \log(n/\beta_n) \left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1\right), \beta_n^{1/2} \left(\frac{N^{\hat{s}_n}(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (W_1, W_2 \wedge 0)$$

where (W_1, W_2) is $N(\mathbf{0}, \Sigma_{\mathbf{W}})$-distributed with $\Sigma_{\mathbf{W}} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$.
(*ii*) For Uniform(0,1)-distribution G (so $G(x)$ is in Case III for maxima), and positive constants $(\beta_n)_{n\geq 1}$ and $(d_n)_{n\geq 1}$ satisfying $\beta_n/n \to 0$ and $\beta_n^{3/2}/n \to \infty$, with $d_n = \beta_n(1 - \beta_n/(2n))$ (and $\beta_n = d_n(1 - (d_n/(2n)) + o(d_n/n))$), and $\rho_n = 1 - (\beta_n/n)$, Theorem 2.1 gives

$$\left(\beta_n^{1/2} (n/\beta_n) \left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1\right), \beta_n^{1/2} \left(\frac{N^{\hat{s}_n}(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (W_1, W_2 \wedge 0)$$

where (W_1, W_2) is $N(\mathbf{0}, \Sigma_{\mathbf{W}})$-distributed with $\Sigma_{\mathbf{W}} = \begin{pmatrix} 1/3 & -1/2 \\ -1/2 & 1 \end{pmatrix}$.
(*iii*) Let $G(x) = 1 - x^{-1/a}$ for $x \geq 1$, with $0 < a < 1/2$ (so $G(x)$ is in Case II for maxima with index $\alpha > 2$). For positive constants $(\beta_n)_{n\geq 1}$ and $(d_n)_{n\geq 1}$ satisfying $d_n/n \to 0$ and $d_n/n^a \to \infty$, and $\beta_n = ((1-a)d_n n^{-a})^{1/(1-a)}$, with $\rho_n = (n/\beta_n)^a$, Theorem 2.1 gives

$$\left(\beta_n^{1/2} \left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1\right), \beta_n^{1/2} \left(\frac{N^{\hat{s}_n}(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (W_1, W_2 + ((1-a)(W_1 - W_2) \wedge 0))$$

where (W_1, W_2) is $N(\mathbf{0}, \Sigma_{\mathbf{W}})$-distributed with $\Sigma_{\mathbf{W}}$ the left-upper 2×2 matrix as in the corresponding part of Theorem 2.1.
(*iv*) For $G(x)$ as in (*iii*) with $a = 1/2$ (so $G(x)$ is in Case II for maxima with index $\alpha = 2$), and positive constants $(\beta_n)_{n\geq 1}$ and $(d_n)_{n\geq 1}$ satisfying $d_n/n \to 0$ and $d_n/\sqrt{n} \to \infty$, and $\beta_n = d_n^2/(4n)$ with $\rho_n = (n/\beta_n)^{1/2}$, Theorem 2.1 gives

$$\left(\left(\frac{\beta_n}{\log\,\beta_n}\right)^{1/2} \left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1\right), \left(\frac{\beta_n}{\log\,\beta_n}\right)^{1/2} \left(\frac{N^{\hat{s}_n}(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (W_1, W_3 \wedge 0)$$

where (W_1, W_3) is $N(\mathbf{0}, \Sigma_{\mathbf{W}})$-distributed with $\Sigma_{\mathbf{W}} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/4 \end{pmatrix}$.

The following corollary is an immediate consequence of Theorem 2.1.

Corollary 2.6 *Let $G(x)$ be in Case I, II with $0 < a \leq 1/2$, or III for maxima with $G(0-) = 0$ and $0 < r_G \leq \infty$, and let positive constants $(d_n)_{n\geq 1}$, $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ satisfy* (2) *and* (3) *as in Theorem* 2.1. *Then*

(i) the sequence of stopped threshold policies $(\hat{s}_n)_{n\geq 1}$ is a consistent approximator of τ^e;

(ii) $\hat{N}^e(d_n)/N^{\hat{s}_n}(d_n) \to 1$ in probability as $n \to \infty$; and

(iii) for $G(x)$ in Case II with $0 < a \leq 1/2$,

$$\eta_n^{-1}(\hat{N}^e(d_n) - N^{\hat{s}_n}(d_n)) \Rightarrow \hat{\mathcal{N}}^e - \hat{\mathcal{N}}^s$$

with $\eta_n = \beta_n^{1/2}$ for $0 < a < 1/2$, and $= (\beta_n \log\, \beta_n)^{1/2}$ for $a = 1/2$, and

$$(N^{\hat{s}_n}(d_n) - \beta_n)/(\hat{N}^e(d_n) - \beta_n) \Rightarrow \hat{\mathcal{N}}^s/\hat{\mathcal{N}}^e$$

where $\hat{\mathcal{N}}^e$ and $\hat{\mathcal{N}}^s$ are given in Theorem 2.1.

3 Stable convergence

When the d.f. G has a 'fatter tail', that is, for G in Case II for maxima with index $0 < \alpha < 2$, the asymptotic distributional convergence of $\hat{N}^e(d_n)$ and $N^{\hat{s}_n}(d_n)$ toward stable distributions is analyzed in this section by using variations of the techniques of proof of Theorem 2.1. For G in Case II for maxima with $1/2 < a < \infty$, $a = 1/\alpha$, and for positive constants $(d_n)_{n\geq 1}$, define constants $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ by

$$d_n = \begin{cases} n\int_0^{\beta_n/n} G^{\leftarrow}(1-s)ds & \text{for} \quad 1/2 < a < 1 \\ n\int_{1/n}^{\beta_n/n} G^{\leftarrow}(1-s)ds & \text{for} \quad 1 \leq a < \infty \end{cases} \quad \text{and } \rho_n = G^{\leftarrow}(1-(\beta_n/n)). \tag{5}$$

For $1/2 < a < \infty$, let $\Delta(a)$ denote a completely asymmetric stable r.v. with index $\alpha = 1/a$, that is, with characteristic function

$$\phi_{\alpha,\beta,\gamma,\theta}(t) = \exp\{i\theta t - \gamma|t|^{\alpha}(1 - i\beta \text{sgn}(t)\omega(t))\}$$

where function $\omega(t) = \omega(t, \alpha)$, location parameter $\theta = \theta(\alpha)$ and scale parameter $\gamma = \gamma(\alpha)$ are given by $\omega(t) = \tan(\pi\alpha/2)$ if $\alpha \neq 1$, and $= -(2/\pi)\log|t|$ if $\alpha = 1$; $\theta = 0$ if $\alpha \neq 1$, and $= \int_0^{\infty}\left(\frac{\sin x}{x^2} - \frac{1}{x(1+x)}\right)dx$ if $\alpha = 1$; and $\gamma = \Gamma(1-\alpha)\cos(\pi\alpha/2)$ if $0 < \alpha < 1$, $= \pi/2$ if $\alpha = 1$, and $= \alpha(\alpha-1)^{-1}\Gamma(2-\alpha)\cos(\pi\alpha/2)$ if $1 < \alpha < 2$, and skewness parameter $\beta = 1$. (The r.v. $\Delta(a)$ can also be expressed in terms of a rate one Poisson process $N(t)$, $t \geq 0$: for example,

$\Delta(a) = \int_0^\infty N(t)t^{-a-1}dt$ if $a > 1$; see Csörgő et al. [8, page 104].) The assumption in this section which is the counterpart of condition $(3i)$ and (3b) in Section 2 is the following:

for d.f. $G(x)$ in Case II with $1/2 < a < \infty$,

$$\beta_n^{1-a}\left(\frac{1-G(\rho_n-)}{\beta_n/n} - 1\right) \quad \to \quad 0 \text{ as } n \to \infty. \tag{6}$$

Again, for a continuous d.f. $G(x)$, condition (6) is clearly satisfied. In the theorems of this section, Condition (6) is an additional assumption if $1/2 < a < 1$; for sequences $(\beta_n)_{n\geq 1}$ in the theorems, Condition (6) follows immediately from d.f. $G(x)$ being in Case II for maxima if $a \geq 1$.

Theorem 3.1 *Let $G(x)$ be in Case II for maxima with $1/2 < a \leq 1$ and with $G(0-) = 0$, and let the positive constants $(d_n)_{n\geq 1}$, $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ satisfy (5). Assume that $\beta_n \to \infty$ and $\beta_n/n \to 0$ as $n \to \infty$ and that (6) holds. Then there exist positive constants $(\alpha_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq 1}$ for which*

$$\left(\alpha_n^{-1}\left(\hat{N}^e(d_n) - \beta_n\right), \gamma_n^{-1}\left(N^{\hat{s}_n}(d_n) - \beta_n\right)\right) \Rightarrow \left(\hat{\mathcal{N}}^e, \hat{\mathcal{N}}^s\right)$$

where $\left(\hat{\mathcal{N}}^e, \hat{\mathcal{N}}^s\right) = (R_1, R_2 \wedge 0)$ for some nondegenerate r.v.'s R_1 and R_2.

The constants $(\alpha_n)_{n\geq 1}$, $(\gamma_n)_{n\geq 1}$ satisfy $\alpha_n \approx \beta_n^a$ for $1/2 < a \leq 1$, and $\gamma_n \approx \beta_n^a$ for $1/2 < a < 1$ and $\approx \beta_n/\log\beta_n$ for $a = 1$; and the pair (R_1, R_2) is given by

$$(R_1, R_2) = \begin{cases} (-\Delta(a), -(1-a)\Delta(a)) & \textit{for} \quad 1/2 < a < 1 \\ (\exp(-\Delta(1)) - 1, -\Delta(1)) & \textit{for} \quad a = 1. \end{cases}$$

The following constants and functions are used to identify norming constants in the proof of Theorem 3.1:

$$A_n^e = n^a L(1/n) \text{ and } B_n(s) = \begin{cases} n\int_0^s G^{\leftarrow}(1-u)du & \text{for} \quad 1/2 < a < 1 \\ n\int_{1/n}^s G^{\leftarrow}(1-u)du & \text{for} \quad a = 1 \\ 0 & \text{for} \quad 1 < a. \end{cases}$$

¿From Theorem 3 of Csörgő, Horváth and Mason [10], it follows that for $G(x)$ in Case II for maxima with $1/2 < a < \infty$, and for any constants $(k_n)_{n\geq 1}$ with $1 \leq k_n \leq n$ and $k_n \approx \beta_n$

$$\begin{aligned} &(A_n^e)^{-1}\left(\sum_{i=1}^{k_n} Y_{n-i+1,n} - B_n(k_n/n)\right) \\ &= \quad (A_n^e)^{-1}\left(\sum_{i=1}^{n} Y_i - B_n(1)\right) + o_P(1) \stackrel{\mathrm{d}}{=} \Delta(a) + o_P(1) \end{aligned} \tag{7}$$

The statement and proof of the following lemma should be compared to that of Theorem 3 of Csörgő et al. [10] and to Corollary 2 of Csörgő, Haeusler and Mason [11]; in particular, contrast (7) and (8) in the case of $k_n = j_n = [nc]$ for $0 < c < 1$.

Lemma 3.2 *Let $G(x)$ be in Case II for maxima with $1/2 < a < \infty$ and $G(0-) = 0$; and let positive constants $(d_n)_{n\geq 1}$, $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ satisfy (5) with $\beta_n \to \infty$ and $\beta_n/n \to 0$ as $n \to \infty$. Let $\tau_n := 1 - G(\rho_n-)$ for $n \geq 1$. Then for any integers $(j_n)_{n\geq 1}$ with $1 \leq j_n \leq n$ and $j_n \approx cn$ for some constant $0 < c \leq 1$, it follows that*

$$(A_n^e)^{-1}\left(\sum_{i=1}^{j_n} Y_i I(Y_i \geq \rho_n) - (j_n/n)B_n(\tau_n)\right) \Rightarrow \begin{cases} c^a\Delta(a) & \text{for} \quad a \neq 1 \\ c(\Delta(1) + \log\, c) & \text{for} \quad a = 1. \end{cases} \tag{8}$$

In particular, for $j_n \approx n$,

$$\begin{aligned} &(A_n^e)^{-1}\left(\sum_{i=1}^{j_n} Y_i I(Y_i \geq \rho_n) - (j_n/n)B_n(\tau_n)\right) \qquad (9)\\ &= \quad (A_n^e)^{-1}\left(\sum_{i=1}^{n} Y_i - B_n(1)\right) + o_P(1) \stackrel{\mathrm{d}}{=} \Delta(a) + o_P(1). \end{aligned}$$

Proof. For the proof of (8), first write

$$\begin{aligned} &(A_n^e)^{-1}\left(\sum_{i=1}^{j_n} Y_i I(Y_i \geq \rho_n) - (j_n/n)B_n(\tau_n)\right) \\ &= \quad o(1) + \left(\frac{\sum_{i=1}^{j_n} Y_i - (j_n/n)B_n(1)}{j_n^a L(1/j_n)}\right)\left(\frac{j_n^a L(1/j_n)}{n^a L(1/n)}\right) \\ &\quad - \left((A_n^e)^{-1} n^{1/2}\tau_n^{(1/2)-a} L(\tau_n)\right)\left(\frac{\sum_{i=1}^{j_n} Y_i I(Y_i < \rho_n) - j_n\int_{\tau_n}^1 G^{\leftarrow}(1-s)ds}{n^{1/2}\tau_n^{(1/2)-a}L(\tau_n)}\right). \end{aligned}$$

Observe that the second expression on the right converges in distribution to $c^a\Delta(a)$ for $a \neq 1$ and to $c(\Delta(1) + \log\, c)$ for $a = 1$; and in the third expression

$$(A_n^e)^{-1} n^{1/2}\tau_n^{(1/2)-a} L(\tau_n) = o(1)$$

(e.g., use Lemma 2 in [10]) and the quotient in parentheses converges to a $N(0, \sigma^2)$-distributed r.v., with σ^2 finite, since it equals

$$c^{1/2}\left(\tau_n^{-1/2} U_{j_n}(1-\tau_n) - \frac{\int_{1/n}^{1-\tau_n} U_{j_n}(s)dG^{\leftarrow}(s)}{\tau_n^{1/2} G^{\leftarrow}(1-\tau_n)}\right) + o_P(1).$$

To prove (9), use an argument analogous to that used for (8), together with the observation that $(A_n^e)^{-1}\left(\sum_{i=j_n+1}^{n} Y_i - \left(\frac{n-j_n}{n}\right)B_n(1)\right) = o_P(1)$. □

If, in addition to the hypotheses of Lemma 3.2, it is true that d.f. satisfies condition (6), then the conclusion and proof of Lemma 3.2 hold with τ_n replaced by β_n/n.

Proof of Theorem 3.1. The verifications in the cases $1/2 < a < 1$ and $a = 1$ are analogous, so arguments are given here for $1/2 < a < 1$, and not for $a = 1$. First, let $(k_n)_{n\geq 1}$ and $(j_n)_{n\geq 1}$ be given by $k_n = \lceil \beta_n + \alpha_n \mu \rceil$ and $j_n = \lceil n(1 + \gamma_n \beta_n^{-1} \eta) \rceil$ for $-\infty < \mu < \infty$ and $-\infty < \eta < 0$, and use results on slowly varying functions (see the Appendix) to obtain

$$(A_n^e)^{-1}(d_n - B_n(k_n/n)) = -\mu + o(1) \text{ and } (A_n^e)^{-1}(d_n - (j_n/n)B_n(\beta_n/n)) = -(1-a)^{-1}\eta + o(1).$$

Next, use that $\hat{\nu}_n(d_n)/n \to 1$ in probability (from (8)) to obtain that

$$\gamma_n^{-1} \sum_{i=1}^{\hat{\nu}_n(d_n)} (I(Y_i \geq \rho_n) - (1 - G(\rho_n -))) = o_P(1).$$

¿From these results, (7) and (9), and (6), obtain for $-\infty < \mu < \infty$ and $-\infty < \eta < 0$ that

$$\begin{aligned}
&P(\alpha_n^{-1}(\hat{N}^e(d_n) - \beta_n) \leq \mu, \gamma_n^{-1}(N^{\hat{s}_n}(d_n) - \beta_n) \leq \eta) \\
&= P(\hat{N}^e(d_n) \leq \alpha_n \mu + \beta_n, \hat{\nu}_n(d_n) \leq n(1 + \beta_n^{a-1}\eta)) + o(1) \\
&= P\left(\sum_{i=1}^{k_n} Y_{n-i+1,n} > d_n, \sum_{i=1}^{j_n} Y_i I(Y_i \geq \rho_n) > d_n \right) + o(1) \\
&= P\Bigg((A_n^e)^{-1} \left(\sum_{i=1}^{n} Y_i - n \int_0^1 G^{\leftarrow}(1-u)du \right) > -\mu, \\
&\qquad (A_n^e)^{-1} \left(\sum_{i=1}^{n} Y_i - n \int_0^1 G^{\leftarrow}(1-u)du \right) > -(1-a)^{-1}\eta \Bigg) + o(1) \\
&= P\Bigg((-1)(A_n^e)^{-1} \left(\sum_{i=1}^{n} Y_i - n \int_0^1 G^{\leftarrow}(1-u)du \right) < \mu, \\
&\qquad -(1-a)(A_n^e)^{-1} \left(\sum_{i=1}^{n} Y_i - n \int_0^1 G^{\leftarrow}(1-u)du \right) < \eta \Bigg) + o(1)
\end{aligned}$$

and the conclusion follows. □

Observe that in the case $a = 1$ of Theorem 3.1, the location parameter is β_n for both $\hat{N}^e(d_n)$ and $N^{\hat{s}_n}(d_n)$; and the scaling constants are β_n and $\beta_n / \log \beta_n$ for $\hat{N}^e(d_n)$ and $N^{\hat{s}_n}(d_n)$ respectively (with $\beta_n >> \beta_n / \log \beta_n$ in this case). Note also that $\beta_n / n \to 0$, but in this case, $d_n/n \approx (\log \beta_n) L(\beta_n/n)$. Finally, in this case the distributional convergence of $\hat{N}^e(d_n)$ can be rewritten as $\beta_n^{-1} \hat{N}^e(d_n) \Rightarrow \exp(-\Delta(1))$.

Example 3.3 Let $G(x) = 1 - x^{-1/a}$ for $x \geq 1$, with $1/2 < a \leq 1$, so $G(x)$ is in Case II for maxima with index $1 \leq \alpha < 2$.

(i) For $1/2 < a < 1$, and positive constants $(d_n)_{n\geq 1}$ satisfying $d_n/n \to 0$ and $d_n/n^a \to \infty$, Theorem 3.1 gives

$$\left(\beta_n^{1-a}\left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1\right), \beta_n^{1-a}\left(\frac{N^{\hat{s}_n}(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (-\Delta(a), (-(1-a)\Delta(a)) \wedge 0)$$

where $\Delta(a)$ is a r.v. with stable distribution of index $\alpha = 1/a$. Note that β_n can be written as $\beta_n = ((1-a)d_n n^{-a})^{1/(1-a)}$.

(ii) For $a = 1$ and positive constants $(d_n)_{n\geq 1}$ satisfying $d_n/n \to \infty$ and $(d_n/n) - \log n \to -\infty$, $\beta_n = \exp(d_n/n)$ (see (5)) and Theorem 3.1 gives

$$\left(\frac{\hat{N}^e(d_n)}{\beta_n} - 1, (\log \beta_n)\left(\frac{N^{\hat{s}_n}(d_n)}{\beta_n} - 1\right)\right) \Rightarrow (\exp(-\Delta(1)) - 1, (-\Delta(1)) \wedge 0)$$

where $\Delta(1)$ is a r.v. with stable distribution of index $\alpha = 1$.

The next corollary is the 'stable' analogue of Corollary 2.6.

Corollary 3.4 *Let $G(x)$ be in Case II for maxima with $1/2 < a \leq 1$ and $G(0-) = 0$, and let positive constants $(d_n)_{n\geq 1}$, $(\beta_n)_{n\geq 1}$ and $(\rho_n)_{n\geq 1}$ satisfy the hypotheses of Theorem* 3.1. *Then*
(i) *the sequence of stopped threshold policies $(\hat{s}_n)_{n\geq 1}$ is a consistent approximator of τ^e;*
(ii) *for $G(x)$ in Case II with $1/2 < a < 1$, $\hat{N}^e(d_n)/N^{\hat{s}_n}(d_n) \to 1$ in probability as $n \to \infty$; and*
(iii) *for $G(x)$ in Case II with $1/2 < a < 1$,*

$$\begin{aligned} \beta_n^{-a}(\hat{N}^e(d_n) - N^{\hat{s}_n}(d_n)) &\Rightarrow \hat{\mathcal{N}}^e - \hat{\mathcal{N}}^s \text{ and} \\ (N^{\hat{s}_n}(d_n) - \beta_n)/(\hat{N}^e(d_n) - \beta_n) &\Rightarrow \hat{\mathcal{N}}^s/\hat{\mathcal{N}}^e \end{aligned}$$

where $\hat{\mathcal{N}}^e$ and $\hat{\mathcal{N}}^s$ are given as in Theorem 3.1.

For distributions $G(x)$ in Case II with $1 < a < \infty$ the following two results demonstrate that, in contrast to Cases I, II with $0 < a \leq 1$ and III, the stopped threshold policy $\hat{s}_n$ with threshold $\rho_n = G^{\leftarrow}(1 - (\beta_n/n))$ is *not* a consistent approximator of the off-line largest-fit strategy τ^e.

Theorem 3.5 *Let $G(x)$ be in Case II with $1 < a < \infty$ and with $G(0-) = 0$, and let positive constants $(d_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ satisfy* (5). *Assume that $\beta_n \to \infty$ as $n \to \infty$. Then*

$$\hat{N}^e(d_n) \Rightarrow \hat{\mathcal{N}}^e := \sup\left\{k \geq 1 : \sum_{j=1}^{k} S_j^{-a} \leq 1/(a-1)\right\}$$

if this set is nonempty and $:= 0$ *otherwise, where* $S_j = T_1 + \cdots + T_j$ *and* $T_1, T_2, \cdots$ *are standard exponentially distributed r.v.'s.*

Proof. The theorem follows immediately from the convergence $d_n/(n^a L(1/n)) \to 1/(a-1)$ which can be obtained from results on slowly varying functions (see the Appendix), and the following result which can be found in Csörgő and Mason [9, page 974]:

$$\frac{\sum_{j=1}^{k} Y_{n-j+1,n}}{n^a L(1/n)} \Rightarrow \sum_{j=1}^{k} S_j^{-a}, \text{ for fixed } k \geq 1,$$

with $\{S_j\}$ as in the assertion of the theorem. □

Theorem 3.6 *Let* $G(x)$ *be in Case II with* $1 < a < \infty$ *and with* $G(0-) = 0$*, and let the positive constants* $(d_n)_{n\geq 1}$*,* $(\beta_n)_{n\geq 1}$ *and* $(\rho_n)_{n\geq 1}$ *satisfy* (5)*. Assume that* $\beta_n \to \infty$ *and* $\beta_n/n \to 0$ *as* $n \to \infty$*. Then*

$$\beta_n^{-1} N^{\hat{s}_n}(d_n) \Rightarrow \hat{\mathcal{N}}^s$$

where $\hat{\mathcal{N}}^s = \Big(((a-1)\Delta(a))^{-1/a} I(\Delta(a) > 0)\Big) \wedge 1$ *and* $\Delta(a)$ *is a r.v. with stable distribution of index* $\alpha = 1/a$*.*

Proof. First, use Chebychev's inequality to obtain that for each $0 < \gamma \leq 1$,

$$[\gamma n]^{-1} \sum_{j=1}^{[\gamma n]} I(Y_j \geq \rho_n) n \beta_n^{-1} \to 1 \text{ in probability as } n \to \infty. \tag{10}$$

Second, from properties of slowly varying functions, it follows that $d_n/(n^a L(1/n)) = (a-1)^{-1} + o(1)$, and hence from the definition of $\hat{\nu}_n(d_n)$, Lemma 3.2, and the assumption that $\beta_n \to \infty$ and $\beta_n/n \to 0$, obtain

$$\hat{\nu}_n(d_n)/n \Rightarrow \hat{\mathcal{N}}^s \tag{11}$$

where $\hat{\mathcal{N}}^s$ is defined above.

Now, one can use (10) and (11) and an argument analogous to that used to prove Theorem 9.4.1 in [4], Theorem 1 in [13], or Lemma 1 in [1], to obtain

$$(\hat{\nu}_n(d_n))^{-1} \sum_{j=1}^{\hat{\nu}_n(d_n)} I(Y_j \geq \rho_n) n \beta_n^{-1} \to 1 \text{ in probability as } n \to \infty.$$

Finally, use this convergence result, (11) and the representation

$$\beta_n^{-1} N^{\hat{s}_n}(d_n) = \left((\hat{\nu}_n(d_n))^{-1} \sum_{j=1}^{\hat{\nu}_n(d_n)} I(Y_j \geq \rho_n) n \beta_n^{-1} \right) (\hat{\nu}_n(d_n)/n)$$

to show that $\beta_n^{-1} N^{\hat{s}_n}(d_n) = (\hat{\nu}_n(d_n)/n) + o_P(1)$; and the conclusion follows. □

4 Large capacities

The paper is concluded with the analogue of Theorems 2.1 and 3.1 in the 'large capacity' setting, in which it is assumed that the positive capacities $(d_n)_{n\geq 1}$ satisfy $n^{-1/2}(d_n - n\theta) \to 0$ as $n \to \infty$, for some $0 < \theta < EY_1$. Define constants τ and $\rho_n = \rho$ by $\theta = \int_0^\tau G^{\leftarrow}(1-s)ds$ and $\rho = G^{\leftarrow}(1-\tau)$. The proof of Theorem 4.1 goes along the same lines as the proof of Theorem 2.1 using standard approximation arguments, for example, given in Csörgő et al. [7], and is therefore omitted here.

Theorem 4.1 *If $G(x)$ is continuous and strictly increasing on its support, $G(0-) = 0$ and $EY_1^2 < \infty$, then*

$$\left(n^{-1/2}(\hat{N}^e(d_n) - n\tau), n^{-1/2}(N^{\hat{s}_n}(d_n) - n\tau)\right) \Rightarrow (\hat{\mathcal{N}}^e, \hat{\mathcal{N}}^s)$$

where $(\hat{\mathcal{N}}^e, \hat{\mathcal{N}}^s) = (W_1, W_2 + (W_3 \wedge 0))$ and (W_1, W_2, W_3) is $N(\mathbf{0}, \Sigma(\theta))$-distributed. The covariance matrix $\Sigma(\theta)$ is a symmetric matrix $\Sigma(\theta) = (\Sigma(\theta)_{i,j})_{i,j=1,2,3}$ given by

$$\begin{aligned}
(\Sigma(\theta)_{1,j})_{j=1,2,3} &= \left(\sigma^2/\rho^2, (1-\tau)(\tau - (\theta/\rho)), \theta^{-1}\tau((\sigma^2/\rho) - (1-\tau)(\tau\rho - \theta))\right);\\
(\Sigma(\theta)_{2,j})_{j=2,3} &= (\tau(1-\tau), -\tau(1-\tau));\\
\Sigma(\theta)_{3,3} &= \theta^{-2}\tau^2\left(\sigma^2 + 2\rho(1-\tau)\theta - \rho^2\tau(1-\tau)\right); \text{ and}\\
\sigma^2 = \sigma^2(\tau) &= \int_{1-\tau}^1 \int_{1-\tau}^1 (s \wedge t - st)dG^{\leftarrow}(s)dG^{\leftarrow}(t).
\end{aligned}$$

In fact, $W_3 = ((\tau\rho)/\theta)(W_1 - W_2)$.

The following corollary follows immediately.

Corollary 4.2 *If $G(x)$ is continuous and strictly increasing on its support, $G(0-) = 0$ and $EY_1^2 < \infty$, then*
(i) the sequence of stopped threshold policies $(\hat{s}_n)_{n\geq 1}$ is a consistent approximator of τ^e;
(ii) $\hat{N}^e(d_n)/N^{\hat{s}_n}(d_n) \to 1$ in probabiltity as $n \to \infty$;
(iii) $(N^{\hat{s}_n}(d_n) - n\tau)/(\hat{N}^e(d_n) - n\tau) \Rightarrow \hat{\mathcal{N}}^s/\hat{\mathcal{N}}^e$; and
(iv) $n^{-1/2}(\hat{N}^e(d_n) - N^{\hat{s}_n}(d_n)) \Rightarrow \hat{\mathcal{N}}^e - \hat{\mathcal{N}}^s$
where $(\hat{\mathcal{N}}^e, \hat{\mathcal{N}}^s)$ is given in Theorem 4.1.

A Appendix

Recall the following definitions, relations, and results concerned with domains of attraction for maxima. An excellent monograph on extreme value theory is Resnick [14].

Let $Y_1, Y_2, \cdots$ be i.i.d. r.v.'s with d.f. G. $G(x)$ is said to be in the domain of attraction of d.f. $H(x)$ for maxima if there exists constants $(a_n)_{n\geq 1}$, $a_n > 0$ and $(b_n)_{n\geq 1}$, $b_n \in \mathbb{R}$, such that

$$P\left(a_n\left(\max_{1\leq i\leq n} Y_i - b_n\right) \leq x\right) \to H(x) \text{ for all continuity points } x \text{ of } H. \tag{12}$$

In this setting we say G is in Case I, in Case II with index $\alpha > 0$ or in Case III with index $\alpha > 0$, if the limit d.f. H is respectively given by $\Lambda(x) = \exp\left(-e^{-x}\right)$ for $x \in \mathbb{R}$; by $\Phi_\alpha(x) = \exp\left(-x^{-\alpha}\right)$ for $x > 0$; or by $\Psi_\alpha(x) = \exp\left(-(-x)^\alpha\right)$ for $x \leq 0$. In the following paragraphs we list some properties in each of the Cases I, II, and III which are used in this paper.

- If $G(x)$ is in Case I for maxima, then constants $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ are given by

$$b_n = G^{\leftarrow}(1 - 1/n) \text{ and } a_n^{-1} = g(b_n) \text{ for } n \geq 1, \tag{13}$$

  for example with $g(t) = \int_t^{r_G}(1 - G(x))dx/(1 - G(t))$. The function $G^{\leftarrow}(1 - s)$ is slowly varying at zero. An auxiliarly function useful in analysis is the function $c(s)$ defined on $[0, 1]$ by $c(s) := s^{-1}\int_{1-s}^{1}(1 - u)dG^{\leftarrow}(u) = s^{-1}\int_0^s G^{\leftarrow}(1 - u)du - G^{\leftarrow}(1 - s)$. As proven by Lo [12], the function $c(s)$ satisfies the following properties. There exists a finite constant k such that for $0 < s \leq 1/2$, $G^{\leftarrow}(1 - s) = k - c(s) + \int_s^1 u^{-1}c(u)du$. Note also that the function $\phi(s) := \int_0^s G^{\leftarrow}(1-u)du$ can be written as $\phi(s) = s\left(k + \int_s^1 u^{-1}c(u)du\right)$. Moreover, $c(s) > 0$; $c(s)$ is slowly varying at zero; $\lim_{s\downarrow 0} c(s)/G^{\leftarrow}(1 - s) = 0$; and if r_G is finite, then $\lim_{s\downarrow 0} c(s) = 0$.

- If $G(x)$ is in Case II for maxima, with $\alpha > 0$, and $a = 1/\alpha$, then

$$a_n^{-1} = G^{\leftarrow}(1 - 1/n) \text{ and } b_n = 0 \text{ for } n \geq 1, \tag{14}$$

  $r_G = \infty$ and $\lim_{n\to\infty} n(1 - G(a_n^{-1}x)) = x^{-\alpha}$ for $x > 0$. The function $1 - G(x)$ is regularly varying as $x \to \infty$ with index $-\alpha = -1/a$; and $G^{\leftarrow}(1 - s) = s^{-a}L(s)$ where L is slowly varying at zero.

- If $G(x)$ is in Case III for maxima, with $\alpha > 0$, and $a = -1/\alpha$, then

$$a_n^{-1} = r_G - G^{\leftarrow}(1 - 1/n) \text{ and } b_n = r_G \text{ for } n \geq 1, \tag{15}$$

  r_G is finite and $\lim_{n\to\infty} n(1 - G(a_n^{-1}x + r_G)) = (-x)^\alpha$ for $x < 0$. The function $1 - G(r_G - x^{-1})$ is regularly varying as $x \to \infty$ with index $-\alpha = 1/a$; and $G^{\leftarrow}(1 - s) = r_G - s^{-a}L(s)$ where L is slowly varying at zero.

We state two results concerning functions which are slowly varying at zero, which we frequently use in this paper.

- (Karamata's Theorem (see [9, Lemma 1]).) Let $L(x)$ be slowly varying at zero. If $\beta < 1$, then
$$\lim_{s\downarrow 0} \int_0^s u^{-\beta} L(u)du/ \left(s^{1-\beta}L(s)\right) = \frac{1}{1-\beta}.$$
- (Karamata Representation (see e.g. [14, Section 0.4]).) For $0 < t < t_0$, $L(t)$ is slowly varying at zero if and only if $L(t) = c(t)\exp\left(\int_t^{t_0} \varepsilon(u)/u\, du\right)$ for some measurable functions $c : (0, t_0) \to \mathbb{R}_+$ and $\varepsilon : (0, t_0) \to \mathbb{R}$ satisfying $\lim_{t\downarrow 0} c(t) = c_0$ for some constant $c_0 > 0$ and $\lim_{t\to 0} \varepsilon(t) = 0$.

Finally, we give an example of a d.f. $G(x)$ which is in Case II for maxima with $a = 1/3$, and a sequence $(d_n)_{n\geq 1}$ for which condition ($3i$) fails. Other examples illustrating conditions (3b) and (6) can be constructed similarly. Let $G(x) = 1 - [x]^{-3}$ for $x \geq 1$, where $[x]$ is the greatest integer less than or equal to x; thus $G(x) = \sum_{k=1}^{\infty}(1 - k^{-3})I(k \leq x < k+1)$. It is clear that $G(x)$ is a d.f. in Case II for maxima with $\alpha = 3$ and $a = 1/3$. Let $B_k := 1 - k^{-3} = G(k)$ for $k = 1, 2, \cdots$; and observe that $G^{\leftarrow}(w) = \sum_{k=2}^{\infty} kI(B_{k-1} < w \leq B_k)$ for $0 < w < 1$. Let $k_n := [\log n]$ for $n = 1, 2, \cdots$, and let $(d_n)_{n\geq 1}$ be the positive integers satisfying
$$d_n/n = \int_{B_{k_n}}^1 G^{\leftarrow}(u)du = \sum_{l=k_n}^{\infty}(l+1)(B_{l+1} - B_l) = k_n^{-2} + \sum_{l=k_n}^{\infty} l^{-3} \text{ for } n \geq 1.$$
Note that $d_n \approx (3/2)n/(\log n)^2$. These $(d_n)_{n\geq 1}$ were chosen so that the constants $(\beta_n)_{n\geq 1}$ of (2) are given by
$$\beta_n = n(1 - B_{k_n}) = nk_n^{-3} \text{ for } n \geq 1;$$
note that $\beta_n \approx n/(\log n)^3$. Also the constants $(\rho_n)_{n\geq 1}$ of (2) satisfy
$$\rho_n = G^{\leftarrow}(1 - (\beta_n/n)) = G^{\leftarrow}(B_{k_n}) = k_n \text{ for } n \geq 1.$$
Thus, it follows that for this example
$$\beta_n^{1/2}\left(\frac{1 - G(\rho_n-)}{\beta_n/n} - 1\right) = n^{1/2}(1 - B_{k_n})^{1/2}\left(\frac{1 - B_{k_n-1}}{1 - B_{k_n}} - 1\right) \approx 3n^{1/2}k_n^{-5/2} \approx (1/5)(n/\log n)^{1/2},$$
which goes to infinity as $n \to \infty$.

Acknowledgment. The authors are grateful to the School of Mathematics of the Georgia Institute of Technology in Atlanta, the Econometric Institute of the Erasmus University

Rotterdam and the Stieltjes Institute in The Netherlands for reciprocal invitations to pursue this research, and for their hospitality.

The research of F.A. Boshuizen was partially supported by a Fulbright Grant. The research of R.P. Kertz was partially supported by National Science Foundation Grant DMS 92-09586.

References

[1] Azlarov, T.A., Dzamiraev, A.A. and Abdullaev, A.G. (1986) Remarks on limit theorems for sequences of random variables with random index. *Prob. Theory and Math. Stat.* **1**, 99-113.

[2] Boshuizen, F.A. and Kertz, R.P. (1995) Smallest-fit selection of random sizes under a sum constraint: weak convergence comparisons, *preprint.*

[3] Bruss, F.T. and Robertson, J.B. (1991) 'Wald's lemma' for sums of order statistics of i.i.d. random variables. *Adv. Appl. Prob.* **23**, 612-623.

[4] Chow, Y.S. and Teicher, H. (1978) *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, New York.

[5] Coffman, E.G., Flatto, L. and Weber, R.R. (1987) Optimal selection of stochastic intervals under a sums constraint. *Adv. Appl. Prob.* **19**, 454-473.

[6] Csörgő, M., Csörgő, S., Horváth, L. and Mason, D.M. (1986) Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31-85.

[7] Csörgő, M., Csörgő, S. and Horváth, L. (1986) *An Asymptotic Theory for Empirical Reliability and Concentration Processes.* Lecture Notes in Statistics **33**. Springer-Verlag, Berlin.

[8] Csörgő, M., Csörgő, S., Horváth, L. and Mason, D.M. (1986) Normal and stable convergence of integral functions of the empirical distribution function *Ann. Prob.* **14**, 86-118.

[9] Csörgő, S. and Mason, D.M. (1986) The asymptotic distribution of sums of extreme values from a regularly varying distribution. *Ann. Probab.* **14**, 974-983.

[10] Csörgő, S., Horváth, L. and Mason, D.M. (1986) What portion of the sample makes a partial sum asymptotically stable or normalΓ *Probab. Th. Rel. Fields* **72**, 1-16.

[11] Csörgő, S., Haeusler, E. and Mason, D. (1991) The asymptotic distribution of extreme sums. *Ann. Probab.* **19**, 783-811.

[12] Lo, G.S. (1989) A note on the asymptotic normality of extreme values. *J. Statist. Planning and Inference* **22**, 127-136.

[13] Mogyoródi, J. (1971) A remark on limiting distributions for sums of a random number of independent random variables. *Rev. Roum. Math. Pures et Appl.* **16**, 552-557.

[14] Resnick, S. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York.